



Ma, J., Liu, W., Miller, P., & Zhou, H. (2016). An evidential fusion approach for gender profiling. *Information Sciences*, 333, 10-20. <https://doi.org/10.1016/j.ins.2015.11.011>

Peer reviewed version

Link to published version (if available):
[10.1016/j.ins.2015.11.011](https://doi.org/10.1016/j.ins.2015.11.011)

[Link to publication record in Explore Bristol Research](#)
PDF-document

This is the author accepted manuscript (AAM). The final published version (version of record) is available online via Elsevier at <http://www.sciencedirect.com/science/article/pii/S002002551500804X> . Please refer to any applicable terms of use of the publisher.

University of Bristol - Explore Bristol Research

General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available: <http://www.bristol.ac.uk/red/research-policy/pure/user-guides/ebr-terms/>

An Evidential Fusion Approach for Gender Profiling

Jianbing Ma¹, Weiru Liu², Paul Miller², and Huiyu Zhou²

¹ Faculty of Engineering and Computing, Coventry University, Coventry, UK, CV1 5FB

² School of Electronics, Electrical Engineering and Computer Science,

Queen's University Belfast, Belfast BT7 1NN, UK

ac1023@coventry.ac.uk, w.liu@qub.ac.uk,

{p.miller,h.zhou}@ecit.qub.ac.uk

Abstract. CCTV (Closed-Circuit TeleVision) systems are broadly deployed in the present world. To ensure in-time reaction for intelligent surveillance, it is a fundamental task for real-world applications to determine the gender of people of interest. However, normal video algorithms for gender profiling (usually face profiling) have three drawbacks. First, the profiling result is always uncertain. Second, the profiling result is not stable. The degree of certainty usually varies over time, sometimes even to the extent that a male is classified as a female, and vice versa. Third, for a robust profiling result in cases that a person's face is not visible, other features, such as body shape, are required. These algorithms may provide different recognition results - at the very least, they will provide different degrees of certainties. To overcome these problems, in this paper, we introduce an Dempster-Shafer (DS) evidential approach that makes use of profiling results from multiple algorithms over a period of time, in particular, Denoeux's cautious rule is applied for fusing mass functions through time lines. Experiments show that this approach does provide better results than single profiling results and classic fusion results. Furthermore, it is found that if severe mis-classification has occurred at the beginning of the time line, the combination can yield undesirable results. To remedy this weakness, we further propose three extensions to the evidential approach proposed above incorporating notions of time-window, time-attenuation, and time-discounting, respectively. These extensions also applies Denoeux's rule along with time lines and take the DS approach as a special case. Experiments show that these three extensions do provide better results than their predecessor when mis-classifications occur.

Keyword: Gender Profiling; Evidence Theory; Cautious Rule; Time-Window; Time-Attenuation; Time-Discounting

1 Introduction

From the beginning of the 21st century, a massive investment has been established in CCTV technology all over the world, e.g., Florida School Bus Surveillance project [1], Federal Intelligent Transportation System Program in the US [3], the First Glasgow Bus Surveillance [21], Intelligent Surveillance Project [9, 12–14, 17, 18, 16, 15], Airport Corridor Surveillance [2], etc. Currently, in the UK, more than four million CCTV

cameras have been operationally deployed. However, the impact of these CCTV systems on preventing anti-social behaviour and criminal events is not satisfactory. For instance, assaults on passengers in public transportation systems, especially on buses and trains, are still a big problem. Although most of the incidents (also called events), are recorded on video cameras, the systems do not provide desirable responses because the data can hardly be actively analyzed in real-time. That is, CCTV cameras operate in a kind of passive mode. They just collect enormous volumes of data with little further utilization. Therefore, to make this technology more effective, CCTV systems have to be active by introducing real-time analysis of video data and providing security alerts such that undesirable behaviour can be stopped or prevented. This change in CCTV capability will significantly increase the chance that offenders are caught in time which brings great advantage in crime prevention.

A crucial and fundamental requirement for developing an active CCTV system is to find and analyze the threat in the scene automatically, which can occur between individuals and undesirable behaviour between individuals and the environment. Computer vision research on this issue has mainly focused on gender/behaviour/action recognition. Based on statistics from criminology studies, most threats are caused by young adolescent males. Hence, for automatic threat assessment, CCTV systems should be able to provide gender and age information for people appeared in the video. In this paper, we focus on the former.

The most obvious cue in determining a person's gender is the appearance of their face. However, for automatic classifiers this usually requires cooperative subjects who are directly looking at the camera and at close range. For most security scenarios one cannot assume this, as the person's face may not be visible as they are facing away from the camera, or they may be too far away - the resulting low resolution making gender discrimination difficult or impossible. Another obvious cue that can help overcome these issues is that of body shape. However, generally automatic classifiers of body shape are a less reliable indicator of gender than face-based classifiers. Furthermore, for both types of classifiers, the output result always has some degree of uncertainty. Secondly, when such classifiers are applied to video sequences, their output can vary significantly with time - even to the extent that a person's gender is incorrectly classified. Thirdly, the key to a robust solution is to use both face and body shape classifiers. Ideally, we would like to use the face classifier result, provided it is detected, otherwise we should resort to using the body shape result. However, this raises the issue of what to do when the outputs of both classifiers are different. To overcome these problems, an evidential (Dempster-Shafer's (DS) theory of evidence) approach is proposed in this paper that makes use of profiling results from multiple profiling algorithms using different human features (e.g., face, full body) over a period of time, in order to provide robust gender profiling of subjects in video.

Imperfect information frequently occurs in video analytic processes. For example, a person may be classified as male with a certainty of 85% by a gender profiling algorithm. However, this does not imply that the person is female with a 15% certainty, rather, we say that the 15% represents what is unknown about the gender, i.e., we do not know how to distribute the remaining 15% between male and female. From probability theory, this information can only be represented as $p(male) \geq 0.85$ and

$p(\text{female}) \leq 0.15$ (or interval probabilities), which is difficult to use for reasoning. Imperfect information is usually caused by ignorance or unreliability of the information sources. For example, a camera may have a faulty gain control setting, illumination could be poor, or the classifier training set may be unrepresentative. Any, or all, of these can result in imperfect information which cannot be represented by probability measures. On the other hand, such imperfect information can be easily handled using an evidential approach, namely, the Dempster-Shafer theory of evidence.

DS theory [4, 22, 10, 11] is a popular framework to deal with uncertain or incomplete information from multiple sources. This theory is capable of modelling incomplete information through ignorance. For combining different pieces of information, DS theory distinguishes two cases, i.e., whether pieces of information are from distinct, or non-distinct, sources. Many combination rules are proposed for information from distinct sources, among which are the well-known Dempster’s rule [22], Smets’ rule [24], Yager’s rule [29], and Dubois & Prade’s hybrid rule [6], etc. In [5], two combination rules, i.e., the cautious rule and the bold disjunctive rule, for information from non-distinct sources are proposed. Subsequently, we view gender profiling results from the same classifier, e.g. face-based, at different times as being from non-distinct sources. For profiling results from different classifiers, they are naturally considered as being from distinct sources. Therefore, all of the problems mentioned above can be handled within the DS framework.

In this paper, for gender profiling results from the same classifier at different time points, Denoeux’s cautious rule [5] is used to combine them. For profiling results from different classifiers (i.e., face profiling and full body profiling), Dempster’s rule [4, 22] is introduced to combine them. And finally, the pignistic transformation is applied to get the probabilities of the subject being male or female.

However, if severe mis-classification occurs at the beginning of the time line, Denoeux’s rule may yield undesirable results. For instance, if a subject is classified as a female with a certainty degree 0.98, and later on it is classified as a male with certainty degrees from 0.85 to 0.95, then by Denoeux’s cautious rule, it will be always classified as a female. In order to remedy this weakness, in this paper, we propose three extensions on applying Denoeux’s rule through time lines, using notions of time-window, time-attenuation, and time-discounting, respectively¹. In the time-window extension, Denoeux’s rule is applied only for the most recent n frames where n is a pre-given threshold depending on the time length. In the time-attenuation extension, the certainty degree is reduced gradually by time at a pre-defined attenuation factor, and in the time-discounting extension, the certainty degree is discounted between the previous certainty degree and the current one by a discounting factor. Experiments show that these three extensions do provide better results when mis-classifications occur, but they have to pay the price of performing less accurate in other situations than the DS fusion method proposed above. In summary, we can say these three extensions are more robust than their predecessor.

¹ In [25], a similar “forgetting” mechanism idea, called *Markovian requirement*, is discussed such that when data are collected sequentially and time is meaningful, the order with which data are collected should be considered.

The DS theory is not a brand-new theory in the computer vision community. In [27, 20], DS theory is used to improve Kalman/Particle filters. The uncertainty they consider are not on classifications from different algorithms but the positions.

Gender profiling with information from multiple sources is not a new issue in computer vision, either. Shan et al. [23] fused gait and face features for improved gender discrimination using canonical correlation analysis, a powerful tool that is well suited to relating two sets of signals. Wang et al. [28] proposed a face representation scheme in which a face is represented in terms of dense Scale Invariant Feature Transform (d-SIFT) and shape contexts of the face image. In [30], face and gait information is used for gender profiling. However, the fusion method is a very simple one, i.e., $p = 0.5 * p_f + 0.5 * p_g$ where p_f is the probability of a subject being a male regarding its face information and p_g is the profiling probability on its gait information. In [7], multi-view gait information (front-end and back-end) is used for gender profiling. This paper also uses a simple fusion method that just adds the probabilities of the multi-view gait profiling results, when the results are normalized to $[0, 1]$. Since we have compared with the fusion method in [31] which is a better alternative than these two methods, in this paper we do not compare with the two approaches.

The remainder of the paper will be organized as follows. In Section 2, we provide the preliminaries on Dempster-Shafer theory. Subsequently, Section 3 introduces the three extensions of the DS approach. In Section 4, we discuss the difficulties in gender profiling in terms of scenarios. Section 5 provides experimental results which show our extensions perform better than its predecessor and a classic fusion approach as well as single profiling approaches. Finally, we conclude the paper in Section 6.

2 Dempster-Shafer Theory

For readers' convenience, here we recall a few basic definitions and concepts in evidence theory, or Dempster-Shafer's theory of evidence.

There are three important functions in the DS theory: the basic belief assignment (bba), the belief function (Bel) and the plausibility function (Pl). The bba maps the power set to the interval $[0, 1]$, assigning 0 to the empty set and bba values summing up to 1 for all the subsets of the power set.

Formally, we denote Ω as a non-empty, finite set of elements, called *the frame of discernment* as follows: $\Omega = \{w_1, \dots, w_n\}$.

Definition 1 A basic belief assignment is a mapping $m : 2^\Omega \rightarrow [0, 1]$ such that

$$\sum_{A \subseteq \Omega} m(A) = 1.$$

If

$$m(\emptyset) = 0,$$

then m is called a **mass function**.

$m(A)$ represents the basic belief assignment for a particular set A , indicating that the chance that the truth element belongs to the set A , but to no particular subset of

A . That is, the value of $m(A)$ pertains only to the set A rather than any subset of A . Otherwise if we know the truth element belongs to some set B which is a subset of A , then this evidence would contribute to $m(B)$ instead of $m(A)$.

For any bba m , if $m(A) > 0$, then A is called a focal element of m . Let \mathcal{F}_m denote the set of focal elements of m . A mass function with only a focal element Ω is called a *vacuous* mass function.

From a mass function m , the *belief* function (Bel) and *plausibility* function (Pl) can be defined to represent the lower and upper bounds of the beliefs implied by m as follows.

$$\begin{aligned} Bel(A) &= \sum_{B: B \subseteq A} m(B) \text{ and} \\ Pl(A) &= \sum_{C: C \cap A \neq \emptyset} m(C). \end{aligned} \quad (1)$$

Belief functions and plausibility functions can be derived from each other as follows:

$$Pl(\bar{A}) = 1 - Bel(A), \quad (2)$$

where \bar{A} is the complement of A .

In addition, we can compute m from Bel (and hence from Pl by Equation 2) as follows:

$$m(A) = \sum_{B: B \subseteq A} (-1)^{|A-B|} Bel(B) \quad (3)$$

Here $|A - B|$ is the difference of the cardinality between the two sets A and B .

In DS theory, we can combine accumulate information/evidence from multiple sources. Many combination methods were proposed, among which the most well-known is the *Dempster's rule of combination*. Formally, let m_1 and m_2 be two mass functions over Ω . The combination result of m_1 and m_2 is a new mass function m given as follows:

$$m(C) = (m_1 \oplus m_2)(C) = \frac{\sum_{A \cap B = C} m_1(A) m_2(B)}{1 - \sum_{A \cap B = \emptyset} m_1(A) m_2(B)} \quad (4)$$

Note that in real-world applications, very likely sources are not always reliable. To take this into account, in [22], a *discounting* method was proposed that a mass function should be discounted by a discounting rate to adapt the source reliability. Let m be a mass function and r be a discounting rate such that $0 \leq r \leq 1$, the discounted mass function m^r is described as follows:

$$m^r(A) = \begin{cases} (1-r)m(A) & A \subset \Omega \\ r + (1-r)m(\Omega) & A = \Omega \end{cases} \quad (5)$$

Here $r = 0$ indicates that the source is completely reliable and $m^r = m$. If $r = 1$, then it means that the source is totally unreliable.

Definition 2 Let m be a bba on Ω . A *pignistic transformation* of m is a probability distribution P_m over Ω such that $\forall w \in \Omega, P_m(w) = \sum_{A \in \mathcal{F}_m} \frac{1}{|\Omega|} \frac{m(A)}{1-m(\emptyset)}$ where $|A|$ is the cardinality of A .

Let \odot be the conjunctive combination operator (or Smets' operator [24]) for any two bbas m, m' over Ω such that

$$(m \odot m')(C) = \sum_{A \subseteq \Omega, B \subseteq \Omega, A \cap B = C} m(A) m'(B), \forall C \subseteq \Omega. \quad (6)$$

A simple bba m such that $m(A) = x, m(\Omega) = 1 - x$ for some $A \neq \Omega$ will be denoted as A^x . The vacuous bba can thus be noted as A^0 for any $A \subset \Omega$. Note that this notation, i.e., A^x , is a bit different from the one defined in [5] in which A^x in our paper should be denoted as A^{1-x} in [5]. We use this notation instead of Denœux's notation to simplify the description of our approach as can be shown by Definition 4 and Definition 5.

Similarly, for two sets $A, B \subset \Omega, A \neq B$, let $A^x B^y$ denote a bba m such that $m = A^x \odot B^y$ where \odot is the conjunctive combination operator defined in Equation (6).

It is easy to prove that any $m = A^x B^y$ is:

$$m(\emptyset) = xy, m(A) = x(1 - y), m(B) = y(1 - x), m(\Omega) = (1 - x)(1 - y) \quad (7)$$

In addition, when normalized, m in Equation 7 is changed to m' as follows.

$$m'(A) = \frac{x(1 - y)}{1 - xy}, m'(B) = \frac{y(1 - x)}{1 - xy}, m'(\Omega) = \frac{(1 - x)(1 - y)}{1 - xy} \quad (8)$$

A bba m with $m(\Omega) = 0$ is called a *dogmatic* bba. Note that for any bba m such that $m(\Omega) = 0$, the discounted bba m^r with $r > 0$ is a non-dogmatic bba. If r is smaller enough, m^r is a good approximation of m .

It is well-known that a non-dogmatic bba m such that $m(\Omega) > 0$ can be uniquely decomposed into the following form [22, 5]:

$$m = \odot_{A \subset \Omega} A^{x(A)}, x(A) \in [0, 1]. \quad (9)$$

Also note that this decomposition can be extended to a dogmatic bba by discounting it with discounting rate ϵ and letting ϵ tend towards 0 [26, 5].

With this decomposition, the cautious combination rule proposed in [5] is defined as follows.

Definition 3 (*Denœux's Cautious Combination Rule*) Let $m_1 = \odot_{A \subset \Omega} A^{x_1(A)}$ and $m_2 = \odot_{A \subset \Omega} A^{x_2(A)}$ be two bbas, then the combined bba by Denœux's cautious combination rule is $m = \odot_{A \subset \Omega} A^{x(A)}$ such that: $x(A) = \max(x_1(A), x_2(A))$.

Remarks: Recall that a bba A^x here should be written as A^{1-x} in [5]. From the cautious combination rule in [5], the combination result of $\odot_{A \subset \Omega} 1 - A^{x_1(A)}$ and $\odot_{A \subset \Omega} 1 - A^{x_2(A)}$ would be $\odot_{A \subset \Omega} A^{x'(A)}$ such that $x'(A) = \min(1 - x_1(A), 1 - x_2(A))$. Therefore, in our format, we have the combination result $\odot_{A \subset \Omega} A^{x(A)}$ such that $x(A) = 1 - x'(A) = \max(x_1(A), x_2(A))$.

For two bbas $A^{x_1} B^{y_1}$ and $A^{x_2} B^{y_2}$, the cautious combination rule is reduced as follows.

Lemma 1 Let $A^{x_1} B^{y_1}$ and $A^{x_2} B^{y_2}$ be two bbas, then the combined bba by Denœux's cautious combination rule is a bba $A^x B^y$ such that: $x = \max(x_1, x_2), y = \max(y_1, y_2)$.

Also, according to [5], for $m_1 = A^{x_1} B^{y_1}$ and $m_2 = A^{x_2} B^{y_2}$, the combined result by Equation (4) is²

$$m_{12} = A^{x_1+x_2-x_1x_2} B^{y_1+y_2-y_1y_2} \quad (10)$$

3 Gender Recognition Scenario

In this section, we provide a detailed description of a gender profiling scenario, which lends itself naturally to a DS approach.

Figure 1 shows three images taken from a video sequence that has been passed through a video analytic algorithm for gender profiling. In this sequence, a female wearing an overcoat with a hood enters the scene with her back to the camera. She walks around the chair, turning, so that her face becomes visible, and then sits down.

Fig. 1. Three images taken from a video sequence

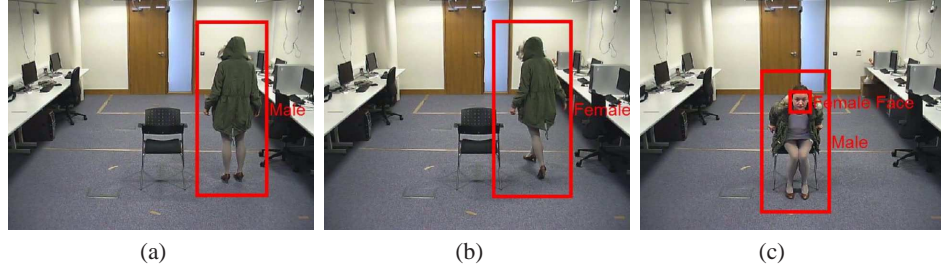


Fig. 1(a) shows that the subject is recognised by the full body shape profiling as a male. Note that her face is not visible. In Fig. 1(b), the subject is classified as female by the full body shape profiling algorithm. In Fig. 1(c), as she sits down, with her face visible, the face profiling algorithm classifies her as female, whilst the full body profiling classifies her as male. Note that the full body profiling algorithm is not as reliable as the face profiling algorithm. Conversely, full body profiling is always possible whilst the face information can be missing. That is why these two profiling algorithms should be considered together. In addition, as full body profiling is not as robust, discount operations should be performed on the algorithm output (cf. Equation (3)). The discount rate is dependent on the video samples and the training efficiency. For every video frame in which a body (face) is detected, gender recognition results are provided. The full body profiling algorithm and the face profiling algorithm, provided a person's face is detected, report their recognition results for every frame of the video, e.g., male with 95% certainty.

For a frame with only a body profiling result, for instance Fig. 1(a), the corresponding mass function m for body profiling will be M^x where M denotes that *the person*

² In [5], the combined result is $m_{12} = A^{x_1x_2} B^{y_1y_2}$, but recall that we use a slightly different notation from [5].

is classified as a male and x is the mass value of $m(\{M\})$. The corresponding mass function for face profiling is $M^0 F^0$ where F denotes that *the person is classified as a female*, or the vacuous mass function. Alternatively, we can refer to this as the vacuous mass function.

Similarly, for a frame with both body profiling and face profiling, for instance Fig. 1(c), the corresponding mass function for body profiling will be M^x (or equivalently $M^x F^0$) and the mass function for face profiling is F^y (or equivalently $M^0 F^y$) where x, y are the corresponding mass values. As time elapses, fusion of bbas by the cautious rule or its three extensions are introduced, as shown by Lemma 1 and Definition 4 and Definition 5. And when it comes to present the final profiling result, we use Dempster's rule to combine the two fused mass functions from the two recognition algorithms, respectively. Namely, for the two bbas $m_1 = M^{x_1} F^{y_1}$ and $m_2 = M^{x_2} F^{y_2}$, it is easy to get that the combined result m_{12} by Dempster's rule is (normalized from the result of Equation 10):

$$\begin{aligned} m_{12}(\{M\}) &= \frac{(x_1 + x_2 - x_1 x_2)(1 - y_1)(1 - y_2)}{1 - (x_1 + x_2 - x_1 x_2)(y_1 + y_2 - y_1 y_2)}, \\ m_{12}(\{F\}) &= \frac{(1 - x_1)(1 - x_2)(y_1 + y_2 - y_1 y_2)}{1 - (x_1 + x_2 - x_1 x_2)(y_1 + y_2 - y_1 y_2)}, \\ m_{12}(\Omega) &= \frac{(1 - x_1)(1 - x_2)(1 - y_1)(1 - y_2)}{1 - (x_1 + x_2 - x_1 x_2)(y_1 + y_2 - y_1 y_2)}. \end{aligned}$$

Finally, we use the pignistic transformation (Def. 2) for the final probabilities. That is, $p(\{M\}) = m_{12}(\{M\}) + m_{12}(\Omega)/2$ and $p(\{F\}) = m_{12}(\{F\}) + m_{12}(\Omega)/2$. Obviously, we will say the subject is a male if $p(\{M\}) > p(\{F\})$, and a female if $p(\{M\}) < p(\{F\})$. In very rare cases that $p(\{M\}) = p(\{F\})$, we cannot know whether it is male or female. Formally, we can write:

$$\text{gender} = \begin{cases} \text{Male} & \text{for } p(\{M\}) > p(\{F\}) \\ \text{Female} & \text{for } p(\{M\}) < p(\{F\}) \\ \text{Unknown} & \text{for } p(\{M\}) = p(\{F\}) \end{cases} \quad (11)$$

The following example illustrates the computation steps.

Example 1 Let us illustrate the approach by a simple scenario with two frames. In the first frame, we have both body profiling (m_b^1) and face profiling (m_f^1) results as $m_b^1 = M^{0.7} F^{0.3}$ and $m_f^1 = M^{0.4} F^{0.6}$. In the second frame, we have the body profiling (m_b^2) result only, where $m_b^2 = M^{0.8} F^{0.2}$.

By Lemma 1, the fusion results by the cautious rule is $m_b = M^{0.8} F^{0.3}$ and $m_f = M^{0.4} F^{0.6}$.

Then by Equation 10, we get $m_{bf} = M^{0.88} F^{0.72}$, which, when normalized, is equivalent to $m_{bf}(\{M\}) = \frac{0.88(1-0.72)}{1-0.88*0.72} = 0.67$, $m_{bf}(\{F\}) = \frac{0.72(1-0.88)}{1-0.88*0.72} = 0.23$, $m_{bf}(\Omega) = \frac{(1-0.88)(1-0.72)}{1-0.88*0.72} = 0.1$.

And finally by pignistic transformation, we get $p(\{M\}) = 0.72$ and $p(\{F\}) = 0.28$ which indicates that the subject is a male.

Actually, we can ease the computation by the following proposition.

Proposition 1 For $m_1 = M^{x_1}F^{y_1}$ and $m_2 = M^{x_2}F^{y_2}$, $p(\{M\}) \geq p(\{F\})$ if and only if $x_1 + x_2 - x_1x_2 \geq y_1 + y_2 - y_1y_2$.

Proof of Proposition 1: It is easy to see

$$\begin{aligned}
& p(\{M\}) \geq p(\{F\}) \\
\iff & m_{12}(\{M\}) \geq m_{12}(\{F\}) \\
\iff & \frac{(x_1 + x_2 - x_1x_2)(1 - y_1)(1 - y_2)}{1 - (x_1 + x_2 - x_1x_2)(y_1 + y_2 - y_1y_2)} \geq \frac{(1 - x_1)(1 - x_2)(y_1 + y_2 - y_1y_2)}{1 - (x_1 + x_2 - x_1x_2)(y_1 + y_2 - y_1y_2)} \\
\iff & (x_1 + x_2 - x_1x_2)(1 - y_1)(1 - y_2) \geq (1 - x_1)(1 - x_2)(y_1 + y_2 - y_1y_2) \\
\iff & (1 - (1 - x_1)(1 - x_2))(1 - y_1)(1 - y_2) \geq (1 - x_1)(1 - x_2)(1 - (1 - y_1)(1 - y_2)) \\
\iff & (1 - y_1)(1 - y_2) \geq (1 - x_1)(1 - x_2) \\
\iff & x_1 + x_2 - x_1x_2 \geq y_1 + y_2 - y_1y_2.
\end{aligned}$$

In the program implementation, we can use Proposition 1 to simplify the computation.

4 Three Extensions

In this section, we discuss three ways to weaken the influence of past information within the framework of DS theory, extended from the aforementioned DS approach using Cautious rule, i.e., a time-window approach, a time-attenuation approach and a time-discounting approach. In the time-window approach, we only consider several recent bbas. In the time-attenuation approaches, we introduce a time-attenuation factor and use this factor to reduce the coefficients of the bbas by time. In the time-discounting approach, we deploy a compromise way commonly used in machine learning areas in which a discounting factor is used to balance between the previous bbas and the current one. We aim to further compare these three approaches to find out the best alternative for remedying the weakness of the Cautious rule when misleading happens in the beginning.

Let \oplus_C be the operator defined by the Cautious rule. We define the three approaches as follows.

Definition 4 (*Time-Window Cautious Combination Rule*) Let $A^{x_1}B^{y_1}, \dots, A^{x_n}B^{y_n}$ be n successive bbas, then the combined bba by Time-Window cautious combination rule of window size t is $m_t = A^{x_{n-t+1}}B^{y_{n-t+1}} \oplus_C \dots \oplus_C A^{x_n}B^{y_n}$.

That is, a time-window combination rule of window size t only combines the recent t bbas. Therefore if a male is mis-classified as a female with a certainty degree 0.98, then after t frames, it will not influence the classification result any more.

Definition 5 (*Time-Attenuation Cautious Combination Rule*) Let $A^{x_1}B^{y_1}, \dots, A^{x_n}B^{y_n}$ be n successive bbas, then the combined bba by Time-Attenuation cautious combination rule of attenuation factor t , $0 < t < 1$, is $m_t = A^{x_1 t^{n-1}}B^{y_1 t^{n-1}} \oplus_C \dots \oplus_C A^{x_n}B^{y_n}$.

That is, in this time-attenuation combination rule of attenuation factor t , the coefficient is reduced by t each time. Hence if a male is mis-classified as a female with a certainty degree 0.98, and hence is represented as $M^0 F^{0.98}$, will be attenuated gradually that it will not affect the cautious combination result for long since $0.98t^n$ will grow smaller when $0 < t < 1$ and n increases.

Definition 6 (*Time-Discounting Combination Rule*) Let $A^{x_1} B^{y_1}, \dots, A^{x_n} B^{y_n}$ be n successive bbas, then the combined bba by Time-Discounting combination rule of attenuation factor t , $0 < t < 1$, is $m_t = A^{x'_n} B^{y'_n}$ where x'_i (resp. y'_i) is defined recursively as: $x'_1 = x_1$, and

$$x'_i = (1 - t)x'_{i-1} + tx_i, i > 1 \text{ (resp. } y'_1 = y_1, \text{ and } y'_i = (1 - t)y'_{i-1} + ty_i, i > 1).$$

That is, in this time-discounting combination rule of discounting factor t , the coefficient is attenuated by the discounting factor t between the previous coefficient and the current one. This kind of discounting is commonly used in machine learning areas [19]. In addition we can see that actually Equation 5 also uses this intuition. This kind of discounting of course reduces the mis-classification quickly since the *history* is always discounted each time.

Here we should notice that the time-window extension and the time-attenuation extension are generalizations to the DS fusion scheme since if the window size equals to the number of frames or the attenuation factor is set to 1, then these two extensions reduce to the DS fusion scheme.

Example 2 Let us illustrate the approach by a simple scenario with four frames, and there is a mis-classification in the first frame. In the first frame, the corresponding both body profiling (m_b^1) and face profiling (m_f^1) results as $m_b^1 = M^{0.6}$ and $m_f^1 = F^{0.9}$ (mis-classification). In the second frame, there is only a body profiling (m_b^2) result which is $m_b^2 = M^{0.7}$. Frame three is associated with body profiling (m_b^3) and face profiling (m_f^3) results as $m_b^3 = F^{0.4}$ and $m_f^3 = M^{0.6}$, and frame four is associated with body profiling (m_b^4) and face profiling (m_f^4) results as $m_b^4 = M^{0.6}$ and $m_f^4 = M^{0.6}$.

By Lemma 1, the fusion results by the cautious rule are $m_b = M^{0.7} F^{0.4}$ and $m_f = M^{0.6} F^{0.9}$.

By Definition 4 with window size 2, the fusion results by the time-window cautious rule are $m_b^W = M^{0.6} F^{0.4}$ and $m_f^W = M^{0.6}$.

By Definition 5 with attenuation factor 0.95, the fusion results by the time-attenuation cautious rule are $m_b^A = M^{0.6} F^{0.38}$ and $m_f^A = M^{0.6} F^{0.77}$.

By Definition 6 with discounting factor 0.9, the fusion results by the time-discounting cautious rule are $m_b^D = M^{0.55} F^{0.04}$ and $m_f^D = M^{0.59}$.

Then by Equation 10, we get $m_{bf} = M^{0.88} F^{0.94}$, which, when normalized, is equivalent to $m_{bf}(\{M\}) = \frac{0.88(1-0.94)}{1-0.88*0.94} = 0.31$, $m_{bf}(\{F\}) = \frac{0.94(1-0.88)}{1-0.88*0.94} = 0.65$, $m_{bf}(\Omega) = \frac{(1-0.88)(1-0.94)}{1-0.88*0.94} = 0.04$. And finally we get $p(\{M\}) = 0.33$ and $p(\{F\}) = 0.67$ which indicates that the subject is a female.

Similarly, we have $m_{bf}^W = M^{0.84}F^{0.4}$, and hence $m_{bf}^W(\{M\}) = \frac{0.84(1-0.4)}{1-0.84*0.4} = 0.76$, $m_{bf}^W(\{F\}) = \frac{0.4(1-0.84)}{1-0.84*0.4} = 0.10$, $m_{bf}^W(\Omega) = \frac{(1-0.84)(1-0.4)}{1-0.84*0.4} = 0.14$ and $p^W(\{M\}) = 0.83$ and $p^W(\{F\}) = 0.17$, which indicates that the subject is a male.

Also, we have $m_{bf}^A = M^{0.88}F^{0.857}$, and hence $m_{bf}^A(\{M\}) = \frac{0.88(1-0.857)}{1-0.88*0.857} = 0.51$, $m_{bf}^A(\{F\}) = \frac{0.857(1-0.88)}{1-0.88*0.857} = 0.42$, $m_{bf}^A(\Omega) = \frac{(1-0.88)(1-0.857)}{1-0.88*0.857} = 0.07$ and $p^A(\{M\}) = 0.55$ and $p^A(\{F\}) = 0.45$ which also supports that the subject is a male.

For the time-discounting approach, we have $m_{bf}^D = M^{0.82}F^{0.04}$, and hence $m_{bf}^D(\{M\}) = \frac{0.82(1-0.04)}{1-0.82*0.04} = 0.81$, $m_{bf}^D(\{F\}) = \frac{0.04(1-0.82)}{1-0.82*0.04} = 0.01$, $m_{bf}^D(\Omega) = \frac{(1-0.82)(1-0.04)}{1-0.82*0.04} = 0.18$ and $p^A(\{M\}) = 0.9$ and $p^A(\{F\}) = 0.1$ which also supports that the subject is a male.

Remarkably, from the definitions and computations above, it is obvious to see that the DS fusion and the three extensions do not bring extra complexity. That is, given the computational complexities of face and body profiling, the complexities of the proposed fusion approaches are merely the sum of those for face and body profiling.

5 Experimental Results

In this section we compare fusion results obtained by a classic approach, a Dempster-Shafer theory approach proposed in Section 3 and three of its extension approaches proposed in Section 4. As there are no benchmark datasets for both body and face profiling statistics, we simulate the output of both body and face classifiers on a sequence containing a male subject (only a single subject). For the body classifier, the probability of any frame being correctly classified as male/female is roughly 60-90%. For the face classifier, only 75% of the available frames are randomly allocated as containing a face. For each of these frames the probability of the frame being correctly classified as being male/female is 85-100%. In both cases the values for $m(\{M\})$ and $m(\{F\})$ are uniformly sampled from the ranges 0.6-0.9 and 0.85-1.0 for the body and face classifiers outputs respectively.

As mentioned before, for gender profiling results from the same classifier at different time points, we use the cautious rule (Lemma 1) to combine them. For profiling results from different classifiers (i.e., face profiling and full body profiling), we use Dempster’s rule (Equation (2)) to combine them. And finally, we apply the pignistic transformation (Def. 2) to get the probabilities of the subject being male or female.

The classic fusion system introduced by Zhou et al. [31] includes a classifier namely “EntropyBoost”, which uses the symmetric Kullback-Leiber divergence to update the learning weights within the standard GentleBoost algorithm. This classifier is able to estimate the gender of a face image through principal component analysis (PCA) eigenvalues and the gender of a full body image through PiHOG feature calculation. Zhou’s method takes the degrees of certainty as probabilities, i.e., they consider the face profiling and the full body profiling output p_f^t and p_b^t indicating the probabilities of faces and full bodies being recognized as males at time t . Then it uses $p_{b,f}^t = c_f^t p_f^t + c_b^t p_b^t$ to calculate the final probability $p_{b,f}^t$ at time t , where c_f^t and c_b^t are the weights of the face and full body profiling at time t , proportional to the feasibility of the two algorithms in

the last twenty frames. As full body profiling is always feasible, suppose face profiling can be applied n times in the last twenty frames, then we have:

$$c_b = \frac{20}{20+n}, c_f = \frac{n}{20+n}.$$

In [31], Zhou et al. compared their fusion schema with several other approaches: (1) a face based scheme with PCA eigenvalues with support vector machine (SVM) (“FACE-PCA”), (2) a full body based approach with HOG features and SVM (“BODY-HOG”), and (3) concatenated HOG features of face/body images with SVM (“CP-FB”), and proved that their algorithm has a significantly higher accuracy than the above three algorithms. Exemplar images of face and/or full body detections and corresponding gender classification outcomes are shown in Figure 2, where Zhou’s algorithm allows one to correctly detect genders on those images.

For this experimental simulation, the performance of the DS and classic (Zhou’s) fusion schemes were characterised by the true positive rate:

$$T_{PR} = \frac{N_{PR}}{N}$$

where N_{PR} is the number of frames in which the gender has been correctly classified and N is the total number of frames in which the body/face is present. According to the training on the sample videos, the discount rate r for the full body profiling is set to 0.3. For comparison, we calculate the T_{PR} value for the body classifier alone, the face classifier, the DS fusion scheme and its three extensions, and the classic fusion scheme.

First we compare the DS approach result and the classic approach result. When applying the methods on the randomly-generated simulation data, the comparison results are presented as follows.

Methods	TotalFrame	N	N_{PR}	T_{PR} (%)
Full Body	3100	3100	1872	60.4
Face	3100	2321	2178	93.8
Classic Method	3100	3100	2658	85.7
DS Approach	3100	3100	3014	97.2

Table 1: Comparison of T_{PR} for body classification, face classification, DS fusion and classic fusion

Note that here the performance of full body or face recognition is generated according to our simulation assumption. Various algorithms may provide different performance values. However, in this paper, we ignore this difference since we are focusing on the comparison of fusion approaches. Indeed the changing of performance values of full body and face recognition does not affect the comparison result between our DS fusion schema and the classic one.

From Table 1, we can see that the DS fusion scheme gives an increase of approximately 13.4% in T_{PR} compared to the classic fusion scheme.

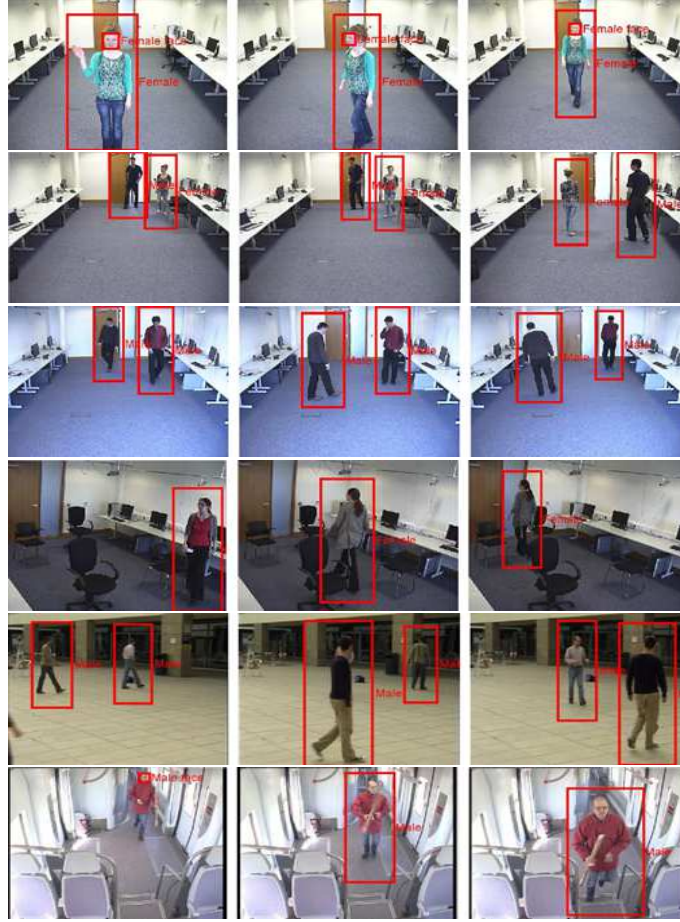


Fig. 2. Examples of correct face/body detection and gender classification in six video sequences using Zhou’s algorithm. Better viewed in color.

Now we show the experimental results on the three extensions. Here we first apply the approaches to 58 simulations each with 50 frames (so there are 2900 total frames), where a mis-classification occurs at the beginning. The comparison results are presented as follows.

Methods	N	N_{PR}	T_{PR} (%)
Full Body	2900	1606	55.4
Face	2159	2002	92.7
Classic Method	2900	2078	71.7
DS Approach	2900	2380	82.1
Time-Attenuation (0.95)	2900	2194	75.7
Time-Attenuation (0.99)	2900	2431	83.8
Time-Discounting (0.85)	2900	2613	90.1
Time-Discounting (0.9)	2900	2471	85.2
Time-Discounting (0.95)	2900	2397	82.7
Time-Window (5)	2900	2586	89.2

Table 2: Comparison of T_{PR} for body classification, face classification, classic fusion, DS fusion and its three extensions - Mis-Classification Cases.

From Table 2, we can see that the three extensions provide better results than the DS fusion scheme, except when the attenuation factor is 0.95. This may be because setting the attenuation factor to 0.95 reduces the certainty degrees too quickly. Also, we can see that the time-window approach is better than the time-attenuation approach, whilst for the time-discounting approach, its precision decreases when the discounting factor increases. This is because in the discounting equation $x'_i = (1-t)x'_{i-1} + tx_i$, the *history* information x'_{i-1} is an integrated value which is more reliable than the single value x_i , so decreasing the contribution of x'_{i-1} will decrease the classification efficiency.

An example simulation result comparing the classic, DS, Time-Attenuation (0.99) and Time-Window (5) approaches (since they are generalizations of the DS fusion scheme) is shown in Figure 3. Figure 3 records data in a single example simulation (50 frames). It shows that when frames increase, the correctly recognized subjects also increases almost linearly.

Here we also show the result comparing the time-discounting approach with different discounting factors in Fig. 4. Figure 4 shows the trend of T_{PR} of the time-discounting approach when the discounting factor increases from 0.05 to 0.95. We can see that when t is increasing from 0.05 to 0.5, T_{PR} is almost not affected, and from $t = 0.5$, it decreases modestly, and from $t = 0.85$, it decreases quickly. This can be explained as when $t \leq 0.5$, the contribution of the *history* is always predominant, so it does not affect much, but when the contribution of the *history* decreases, it indeed influence the classification results. So it justifies that for classification, we should consider the previous history instead of keeping the current classification result only.

Now we apply the approaches to 20 simulations each with 150 frames (so there are 3000 total frames), where we do not assume mis-classification occurred at the beginning. The comparison results are presented as follows.

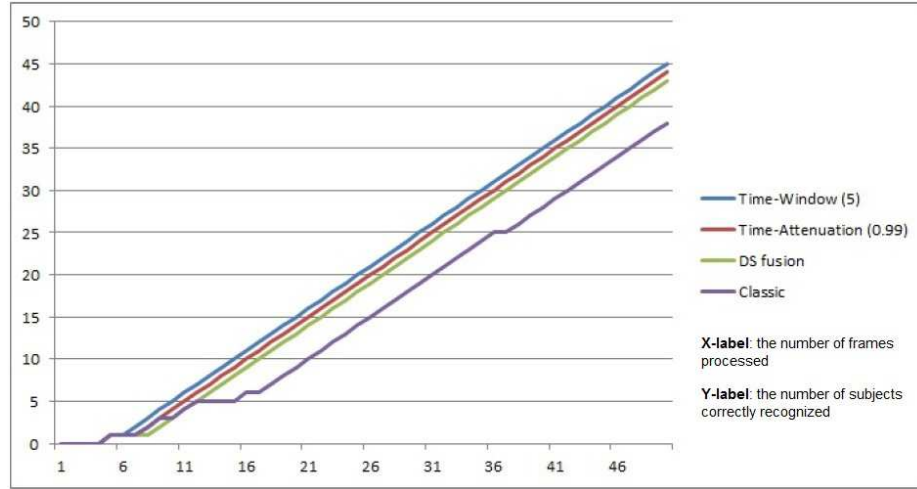


Fig. 3. An Example Simulation



Fig. 4. Time-Discounting with Different Factors

Methods	N	N_{PR}	T_{PR} (%)
Full Body	3000	1792	59.7
Face	2229	2125	95.3
Classic Method	3000	2490	83.0
DS Approach	3000	2899	96.6
Time-Attenuation (0.95)	3000	2126	70.9
Time-Attenuation (0.99)	3000	2401	80.0
Time-Discounting (0.85)	3000	2706	90.2
Time-Discounting (0.9)	3000	2583	86.1
Time-Discounting (0.95)	3000	2501	83.4
Time-Window (5)	3000	2395	79.8
Time-Window (20)	3000	2552	85.1

Table 3: Comparison of T_{PR} for body classification, face classification, classic fusion, DS fusion and its three extensions - General Cases.

From Table 3, we can see that the three extensions perform worse than the DS fusion scheme. This is not surprising since the former do not always hold the highest certainty degree as in the DS fusion scheme. Table 3 also shows that when the attenuation factor or the window size increases, the results improve. Actually, if the window size equals to the number of frames or the attenuation factor is one, then these two extensions (i.e., the time-window extension and the time-attenuation extension) will provide the same results as the DS fusion one. In addition, we can see that the time-discounting approach, in this case, behaves better than the other two extensions. Also, comparing the results in Table 2 and Table 3, we can find that the results of the time-discounting approach are much more stable than those of the other two extensions. The reason is that in the time-discounting approach, the influence of the mis-classification at the beginning disappears more quickly than those of the other two extensions since the discounting factors are large (≈ 1).

6 Conclusion

In this paper, we have proposed how to combine gender profiling classifier results by utilizing DS theory. We have used the cautious rule to combine gender profiling results from the same classifier at different time points and used Dempster’s rule to combine profiling results from different classifiers. Experimental results show that the introduction of the DS theory indeed improves profiling performance. In addition, to deal with mis-classifications occurred at the beginning of the stream, we have proposed two fusion methods by modifying the application of the Cautious rule, i.e., the time-window fusion method and the time-attenuation fusion method. We also have proposed another extension which is the time-discounting fusion method. Experimental results show that these three extensions provide more robust results than other approaches, especially to their predecessor DS fusion scheme.

We have mentioned that there are three problems that a classic gender profiling system should deal with, i.e., uncertain profiling results, unstable results over time for a gender profiling classifier, and different classifiers capturing different features. We have shown that a DS-based approach handles these three issues in a seamless way.

From the experimental results, it suggests that the time-window fusion scheme performs slightly better than the time-attenuation fusion scheme. But we think this conclusion still depends on the choice of attenuation factor, window size and frame size. Also, the time-discounting approach is more stable than the other two extensions.

For future work, we plan to apply the fusion schemes to profiling classifier results generated from large-scale real video sequences. We will combine our work with those of [23, 28] and others in order to improve the performance of our system to handle partial occlusions or crowded situations. In addition, we are also exploiting ideas from knowledge base merging [8].

Acknowledgement This research work is sponsored by the EPSRC projects EP/G034303/1 and EP/H049606/1 (the CSIT project).

References

1. Florida school bus surveillance project, british security industry association. In http://www.bsia.co.uk/LY8VIM18989_action;displaystudy_sectorid;LYCQYL79312_caseid;NFLEN064798.
2. Airport corridor surveillance, ecit, queen's university of belfast. <http://www.csit.qub.ac.uk/Research/ResearchGroups/IntelligentSurveillanceSystems>, 2010.
3. Rita - its research program, us department of transportation. http://www.its.dot.gov/ITS_ROOT2010/its_program/ITSfederal_program.htm, 2010.
4. A. P. Dempster. Upper and lower probabilities induced by a multivalued mapping. *The Annals of Statistics*, 28:325–339, 1967.
5. T. Denœux. Conjunctive and disjunctive combination of belief functions induced by nondistinct bodies of evidence. *Artificial Intelligence*, 172(2-3):234–264, 2008.
6. D. Dubois, H. Prade, and R. Yager. Fuzzy set connectives as combinations of belief structures. *Information Sciences*, 66:245–275, 1992.
7. G. Huang and Y. Wang. Gender classification based on fusion of multi-view gait sequences. In *Procs. of ACCV*, pages 462–471. Springer-Verlag, LNCS 4843, 2007.
8. S. Konieczny and R. Pino-Pérez. On the logic of merging. In A. G. Cohn, L. Schubert, and S. C. Shapiro, editors, *KR'98, Principles of Knowledge Representation and Reasoning*, pages 488–498, San Francisco, California, 1998. Morgan Kaufmann.
9. W. Liu, P. Miller, J. Ma, and W. Yan. Challenges of distributed intelligent surveillance system with heterogenous information. In *Procs. of QRASA*, pages 69–74, Pasadena, California, 2009.
10. J. Ma, W. Liu, D. Dubois, and H. Prade. Revision rules in the theory of evidence. In *Procs. of ICTAI*, pages 295–302, 2010.
11. J. Ma, W. Liu, D. Dubois, and H. Prade. Bridging Jeffrey's rule, AGM revision and Dempster conditioning in the theory of evidence. *International Journal on Artificial Intelligence Tools*, 20(4):691–720, 2011.
12. J. Ma, W. Liu, and P. Miller. Event modelling and reasoning with uncertain information for distributed sensor networks. In *Procs. of SUM*, pages 236–249, Toulouse, France, 2010. Springer.
13. J. Ma, W. Liu, and P. Miller. Belief change with noisy sensing in the situation calculus. In *Procs. of UAI*, 2011.
14. J. Ma, W. Liu, and P. Miller. Handling sequential observations in intelligent surveillance. In *Proceedings of SUM*, pages 547–560, 2011.
15. J. Ma, W. Liu, and P. Miller. Evidential fusion for gender profiling. In *Proceedings of the 6th International Conference on Scaleable Uncertainty Management (SUM'12)*, pages 514–524, 2012.
16. J. Ma, W. Liu, and P. Miller. An evidential improvement for gender profiling. In *Proceedings of 2nd international conference on Belief functions*, pages 29–36, 2012.
17. J. Ma, W. Liu, P. Miller, and W. Yan. Event composition with imperfect information for bus surveillance. In *Procs. of AVSS*, pages 382–387. IEEE Press, 2009.
18. P. Miller, W. Liu, F. Fowler, H. Zhou, J. Shen, J. Ma, J. Zhang, W. Yan, K. McLaughlin, and S. Sezer. Intelligent sensor information system for public transport: To safely go In *Procs. of AVSS*, 2010.
19. T. M. Mitchell. *Machine Learning*. McGraw-Hill Company, ISBN 0-07-042807-7, 1997.
20. T. Reineking. Particle filtering in the Dempster-Shafer theory. *International Journal of Approximate Reasoning*, 52(8):1124–1135, November 2011.
21. Gardiner Security. Glasgow transforms bus security with ip video surveillance. In <http://www.ipusergroup.com/doc-upload/Gardiner-Glasgowbuses.pdf>.

22. G. Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, 1976.
23. C. Shan, S. Gong, and P.W. McOwan. Fusing gait and face cues for human gender recognition. *Neurocomput.*, 71(10-12):1931–1938, 2008.
24. P. Smets. Non-standard logics for automated reasoning. In P. Smets, A. Mamdani, D. Dubois, and H. Prade, editors, *Belief functions*, pages 253–286, 1988.
25. P. Smets. Analyzing the combination of conflicting belief functions. *Information Fusion*, 8(4):387–412, 2007.
26. P. Smets and R. Kennes. The transferable belief model. *Artificial Intelligence*, 66(2):191–234, 1994.
27. P. Smets and B. Ristic. Kalman filter and joint tracking and classification based on belief functions in the tbm framework. *Information Fusion*, 8(1):16–27, 2007.
28. J.-G. Wang, J. Li, W.-Y. Yau, and E. Sung. Boosting dense sift descriptors and shape contexts of face images for gender recognition. In *Procs. of Computer Vision and Pattern Recognition Workshops*, pages 96–102, 2010.
29. R. Yager. On the Dempster-Shafer framework and new combination rules. *Information Sciences*, 41:93–138, 1987.
30. D. Zhang and Y. Wang. Gender recognition based on fusion of face and gait information. In *Proceedings of the Seventh International Conference on Machine Learning and Cybernetics*, pages 62–67, 2008.
31. H. Zhou, P. Miller, J. Zhang, M. Collins, and H. Wang. Gender classification using facial and full body features. In *Technical Report, CSIT, Queen’s University Belfast, UK*, 2011.